



ASSOCIATION FOR CONSUMER RESEARCH

Labovitz School of Business & Economics, University of Minnesota Duluth, 11 E. Superior Street, Suite 210, Duluth, MN 55802

An Extended Paradigm For Measurement Analysis of Marketing Constructs Applicable to Panel Data

Hans Baumgartner, Penn State, US

Jan-Benedict E.M. Sreenkamp, University of Northern Carolina, US

Three extensions of the current scale development paradigm are proposed. First, our model separates stable (trait) sources of substantive variation in a construct from transient (state) sources. Second, we distinguish six types of error based on (a) whether the error is stable or transient, and (b) whether the error affects individual items, subsets of items, or all items in a scale. Third, researchers should make the means of scale items an explicit component of measurement analysis and test for the invariance of item loadings and intercepts for meaningful comparisons. The model is applied to brand loyalty and deal proneness to illustrate the benefits of the proposed procedure.

[to cite]:

Hans Baumgartner and Jan-Benedict E.M. Sreenkamp (2007) , "An Extended Paradigm For Measurement Analysis of Marketing Constructs Applicable to Panel Data", in NA - Advances in Consumer Research Volume 34, eds. Gavan Fitzsimons and Vicki Morwitz, Duluth, MN : Association for Consumer Research, Pages: 537-540.

[url]:

<http://www.acrwebsite.org/volumes/12565/volumes/v34/NA-34>

[copyright notice]:

This work is copyrighted by The Association for Consumer Research. For permission to copy or use this work in whole or in part, please contact the Copyright Clearance Center at <http://www.copyright.com/>.

much less frequently tested. Further, rivals can be conceptually equally plausible. For example, if X =affect, M =cognition, and Y =behavior, there exist supporters of theories which pose $X \rightarrow M \rightarrow Y$, $M \rightarrow X \rightarrow Y$, or $Y \rightarrow M \rightarrow X$, etc. (Breckler, *PsychBull*, 1990).

The regression techniques are no different from the structural equations models in offering no solution to the issue of using all the available degrees of freedom in the decomposition of the variance (the parameter estimates are equivalent). The use of multi-item scales offers no solution to this problem either—additional degrees of freedom are illusory, contributing to the measurement model accuracy, but not to the critical structural model. Ideally, the X , M , Y mediation would be tested in the presence of at least one additional construct, Q , whose role is to serve as an antecedent to X , or a consequence of X , M , or Y .

In the final part of this paper, we extend the SEM models to consider moderated mediation. SAS code is provided for all scenarios.

“Mean-Centering and the Interpretation of ANOVA and Moderated Regression”

Joachim Vosgerau, Carnegie Mellon

Hubert Gatignon, INSEAD

ANOVA is the most prominent statistical analysis in experimental behavioral research. When an interaction of a continuous variable and a categorical variable is hypothesized, it is common practice to dichotomize the continuous scores and analyze the data with an ANOVA. For example, if a researcher wants to test the interaction of familiarity with a product class and gender on the size of the consideration set, the researcher might dichotomize familiarity scores into unfamiliar (score=0) and familiar (score>0), and analyze the data with ANOVA. This practice prevails although dichotomizing continuous variables bears negative consequences such as diminishing statistical power (Irwin and McClelland 2003 *JMR*). The reluctance to analyze data with moderated regression seems to result from the difficulties in interpreting moderated regression results.

Indeed, moderated regression and ANOVA can yield seemingly different effects when applied to the same data. We demonstrate that these differences occur because in ANOVA main effects are estimated at their means, and interaction effects are restricted to be symmetric cross-over effects relative to the means. In contrast, in moderated regression constant effects are estimated at zero, and no specific pattern of interaction is imposed.

In order to make moderated regression and ANOVA effects comparable, some researchers have argued to mean-center variables in moderated regression. We show that mean-centering yields effects that are statistically equivalent to effects from raw variables, but it restricts interaction effects to be symmetric relative to the means of the variables. As a consequence, constant effects in moderated regressions are no longer constant but depend on the specific range of the variables. For example, assume the researcher hypothesizes an interaction of familiarity and gender, such that female consumers who are familiar with the product class include more products in their consideration set than female consumers who are less familiar with the product class. For male consumers no differences with respect to product familiarity are expected. The researcher analyzes the raw data with a moderated regression and finds a significant interaction. In this raw-data analysis, the effect for gender is constant, whether a sample with familiarity scores from 0 to 3 is used or a sample with familiarity scores from 3 to 6. But if the researcher mean-centered familiarity scores, the effect for gender would depend on the specific sample. That is, the researcher would find a different gender effect with the sample of 0 to 3 familiarity scores than with the sample with 3 to 6 scores.

The same problem is shown to hold for main effects in ANOVA. If the researcher dichotomized familiarity scores and analyzed the data with ANOVA, the main effect for gender would depend on the range of the familiarity scores. This undesirable property of range-dependence of main effects in ANOVA is a result of ANOVA constraining interaction effects to be symmetric cross-over effects relative to the mean.

Another interpretational problem with ANOVA concerns the way main effects are estimated. Main effects in ANOVA, as constant effects in mean-centered moderated regression, are computed as the effect when the other variable is at its mean. So, the main effect for familiarity (unfamiliar vs. familiar) is estimated at the mean of gender (female vs. male). But what is the mean of female and male? It clearly makes no sense to talk about the mean of gender. And thus it makes no sense to interpret the main effect of product familiarity that is estimated at the mean of gender. Concluding, in the presence of interaction effects, ANOVA main effects have to be interpreted with caution. Main effects depend on the specific range in which the independent variables are manipulated or measured.

Mean-centering in moderated regression, although it is shown to lead to the undesirable property of range dependence of the constant effects, can nonetheless be beneficial. Mean-centering does not reduce multicollinearity (as some researchers and statistical textbooks claim), but it can help to overcome the arbitrary origin problem of interval scales such as Likert scales. Interval scales have no defined zero point, adding an arbitrary constant to an interval scale does not change its meaning. But adding an arbitrary constant to a continuous variable in a moderated regression changes the effect of the other variable in the moderated regression. So, in moderated regression with an interval scale focal variable, the constant effect of the moderator variable is non-interpretable because of the arbitrary origin of the focal variable. The arbitrary origin of interval scale variables can be eliminated by mean-centering. Mean-centered scores only include deviations from the variable mean, and thus arbitrary constants that are added to the raw variable drop out. However, because mean-centering also changes moderation to a symmetric cross-over interaction effect, caution has to be taken about which variables to mean-center in a moderated regression. We show that when the moderator variable is interval scale, mean-centering this variable allows for interpreting the focal variable's effect as its effect when the moderator is at its mean (the interpretation of the moderator variable's constant effect is not affected by the mean-centering transformation). But when the focal variable is interval scale, mean-centering does not help (as it would change the constant effect of the moderator variable to an effect that varies with the focal variable). In this case, the constant effect of the moderator variable should not be interpreted (the focal variable's effect would not be affected by the mean-centering transformation).

“An Extended Paradigm for Measurement Analysis of Marketing Constructs Applicable to Panel Data”

Hans Baumgartner, Penn State University

Jan-Benedict E.M. Steenkamp, University of Northern Carolina

Although the measurement of marketing and consumer behavior constructs has greatly improved in recent years, we believe that several topics have received insufficient attention in the literature. First, in many areas of research it is necessary to draw a clear distinction between the trait and state aspects of a construct. Although the distinction between states (transitory characteristics of individuals that change with circumstances and over time) and traits (relatively stable dispositions) has led to valuable insights in many areas of behavioral research (e.g., enduring vs. situational

involvement), these two components are not treated separately in conventional measurement models.

Second, measure specificity and other sources of systematic but non-construct-related variation in observed responses are usually confounded with random measurement error in standard measurement analyses. Treating all sources of non-substantive variation as an undifferentiated amalgam of measurement error (a) gives the misleading impression that the causes of measurement error are unknown and unknowable; (b) makes it impossible to get insights into the relative importance of the different sources of error or take ameliorative action; and (c) may lead to improper assessments of reliability and cause other problems as well (e.g., when correlations are corrected for attenuation, the correction may be inappropriate).

Third, the means of observed and latent variables are generally not considered explicitly in measurement analyses because all variables are defined as deviations from their means. This is problematic if item or scale means are to be compared across groups of respondents or over time because the comparisons are only meaningful if certain assumptions of measurement invariance are satisfied.

To overcome these problems, we propose an integrative procedure for conducting measurement analysis that (a) distinguishes temporary and stable components of variance in constructs corresponding to their state and trait aspects; (b) separates systematic sources of non-construct variance from random measurement error; and (c) takes into account the means of the observed variables.

With regard to the first point, we propose a second-order factor specification that enables a sophisticated investigation of the stable and transient components of a construct and avoids the problems associated with assessments of stability via test-retest correlations. For example, the procedure allows researchers to determine what proportion of the variation in a construct is due to states or traits.

With regard to the second point, we propose a new classification of measurement error along two dimensions. The first dimension distinguishes between three types of error depending on how many items are affected: (a) measurement error that is specific to a single item within a scale; (b) measurement error that affects a subset of items within a scale; and (c) measurement error that impacts all the items in a scale. The second dimension distinguishes between sources of error that are restricted to a given occasion of measurement (transient) and those that are consistent over time (stable). We show how the resulting six sources of error (stable scale-wide error, transient scale-wide error, stable item-subset error, transient item-subset error, stable item-specific error, and transient item-specific or random error) can be modeled using a factor-analytic specification and discuss the implications of the model for measurement analysis.

With regard to the third point, the proposed procedure incorporates the item and scale means into the measurement analysis, makes assessments of measurement invariance an explicit component of the scale validation process, and ultimately leads to cross-sectional and longitudinal comparisons of means (as well as relationships between constructs) that are methodologically justified.

An extended application of the model to the constructs of brand loyalty and deal proneness is presented to illustrate the benefits of the proposed procedure. We found that about one quarter of the total variance was systematic (nonrandom) measurement error, most of it stable over time. Moreover, scale reliability using the conventional procedure (such as Cronbach's alpha) was substantially overestimated. This result is probably not unique to the present context because the sources of systematic error will also be present in other scales. In addition, we found that ignoring system-

atic measurement error had a very substantial biasing effect on the results, since the correlation between the substantive constructs of brand loyalty and deal proneness was underestimated by a factor of almost 3 ($-.25$ versus $-.70$). In general, the direction of the bias depends on the sign of the true correlation. Measurement errors are typically positively correlated. If the true correlation between two constructs is negative, it will be underestimated, as in our illustration. If the true correlation is positive, it will be overestimated.

ROUNDTABLE
The Question-Behavior Effect: Current and Future Research
David Sprott, Washington State University, USA

SUMMARY

Asking someone a question about a behavior changes the ultimate performance of that behavior in the future. This *question-behavior effect* was initially demonstrated by Sherman (1980) using self-predictions to influence both socially desirable (making a donation to a charity) and undesirable (singing over the telephone) behaviors. Since that time, two primary groups of researchers have consistently demonstrated the question-behavior effect and have elucidated (at least some of) the theoretical mechanisms underlying the effect.

One group of scholars (who have published research on the mere-measurement effect) have used scaled intention and satisfaction questions to influence most often non-normative consumer-related behaviors (e.g., Chandon, Morwitz, Reinartz, 2004; Dholakia & Morwitz, 2002; Morwitz & Fitzsimons, 2004; Morwitz, Johnson & Schmittlen, 1993; Williams, Fitzsimons, & Block, 2004). These researchers have shown such questions to influence a wide-variety of actions (e.g., first time and repeat purchase of durable and non-durable goods, product choice, transactions with and defection from service providers, flossing, drug and alcohol consumption). The theoretical mechanism expected to underlie these effects is one based on attitude accessibility, whereby the question makes attitudes accessible which guide future performance of focal behaviors.

Another group of researchers (referring to their findings as the self-prophecy effect) have employed dichotomous self-predictions to influence behaviors with clear norms associated with the behavior (e.g., Greenwald, et al., 1987; Spangenberg, 1997; Spangenberg & Greenwald, 1999; Spangenberg, et al., 2003; Sprott, Spangenberg, & Fisher, 2003). In various settings, these researchers have shown that self-predictions can reduce the incidence of non-normative behavior (e.g., cheating on an exam) and increase the rate of normative behavior (e.g., recycling, health club attendance, donating to a charity, gender stereotyping, voting). In contrast to an accessibility perspective, these authors suggest that self-predictions are effective due to the motivation associated with cognitive dissonance. It is argued that the question simultaneously reminds people of prior failures to perform the behavior and norms associated with the target action; any discrepancy between what people have done and what they know they should do produces dissonance, which in turn motivates behavior change.

Both streams of research have demonstrated the importance of the question-behavior effect by reporting findings for real behavior. For example, Spangenberg (1997) found self-predictions to increase health-club attendance rates up to 6 months after the prediction. Similarly, Morwitz et al. (1993) showed that an intention measure influenced car purchase rates up to 6 months after the question was asked. Clearly, the question-behavior effect has significant implications for businesses measuring future behaviors and for organizations that may use questions to alter future behavior. As noted previously, research on the theoretical mechanisms underlying observed effects has documented at least two processes including attitude accessibility and cognitive dissonance. Both mechanisms have received considerable theoretical support in the literature including research demonstrating significant moderators of observed effects. For example, Morwitz and Fitzsimons (2004) demonstrated that prior brand attitudes influenced the magnitude of the question-behavior effect and Sprott, et al. (2003) found that self-

predictions were more effective for those with high (vs. low) social norms. These findings respectively support the attitude accessibility and dissonance interpretations.

Although much has been learned about the question-behavior effect since its introduction over 25 years ago, there is much left to be studied. Research on the theoretical mechanisms driving these effects should continue. Clearly attitude accessibility and cognitive dissonance are likely to underlie at least a portion of the reported findings, still future investigators should examine potential alternate theoretical processes related to these effects that have yet to be explored (e.g., implementation intentions). Other promising areas of research include investigations on the wording of the question and how such changes in wording may influence the magnitude of the effect. Prior research suggests that relatively simple (and outwardly benign) changes to the wording of intention measures can have differential effects in the theory of reasoned action (e.g., Sheppard, Hartwick and Warshaw 1988), similar effects may be possible regarding the question-behavior effect. As noted previously, question-behavior effects have been observed for up to six months after completing the question; such effects are hard to be accounted for via attitude accessibility and cognitive dissonance. Future research therefore needs to investigate further the nature of such long terms effects and the processes by which such effects manifest. Finally, research needs to investigate additional moderators of question behavior effects—most promising moderators would be those that help to clarify the processes underlying the effect.

The purpose of this roundtable is to bring together researchers interested in this area and to discuss unpublished findings and directions for future research. As can be seen from the list of references and those who have committed to attending this roundtable, a fairly substantial number of consumer researchers have investigated (or are at least interested in) the phenomenon. This roundtable is the first of its kind at ACR, although ACR has been the venue for many informal discussions of this topic over the years. Having a formal roundtable is hoped to take these discussions to a new level and stimulate future research into the question-behavior effect.

REFERENCES

- Chandon, P. Morwitz, V.G. & Reinartz, W. J (2004). The short- and long-term effects of measuring intent to repurchase. *Journal of Consumer Research*, 31, 566-572.
- Dholakia, U.M. & Morwitz, V. G. (2002). The scope and persistence of mere-measurement effects: Evidence from a field-study of customer satisfaction measurement. *Journal of Consumer Research*, 29, 159-167.
- Greenwald, A. G., Carnot, C. G., Beach, R., & Young, B. (1987). Increasing voting behavior by asking people if they expect to vote. *Journal of Applied Psychology*, 72 (May), 315-318.
- Morwitz, V. G. & Fitzsimons, G. J. (2004). The mere-measurement effect: Why does measuring intentions change actual behavior? *Journal of Consumer Psychology*, 14 (1&2), 64-74.
- Morwitz, V. G., Johnson, E., & Schmittlen, D. (1993). Does measuring intent change behavior. *Journal of Consumer Research*, 20 (June), 46-51.

- Sheppard, B. H., Hartwick, J., & Warshaw, P. R. (1988). The theory of reasoned action: A meta-analysis of past research with recommendations for modifications and future research. *Journal of Consumer Research*, 15 (3), 325-343.
- Sherman, S. J. (1980). On the self-erasing nature of errors of prediction. *Journal of Personality and Social Psychology*, 39 (August), 211-221.
- Spangenberg, E. R. (1997). Increasing health club attendance through self-prophecy. *Marketing Letters*, 8 (1), 23-32.
- Spangenberg, E. R. & Greenwald, A. G. (1999). Social influence by requesting self-prophecy. *Journal of Consumer Psychology*, 8 (1), 61-89.
- Spangenberg, E. R., Sprott, D. E., Grohmann B., & Smith, R. J. (2003). Mass-communicated prediction requests: Practical application and a cognitive dissonance explanation for self-prophecy. *Journal of Marketing*, 67 (July), 47-62.
- Sprott, David E., Spangenberg, E. R., & Fisher, R. J. (2003). The importance of normative beliefs to the self-prophecy effect. *Journal of Applied Psychology*, 88 (June), 423-31.
- Williams, P., Fitzsimons, G. J. & Block, L. G. (2004). When consumers don't recognize "benign" intentions questions as persuasion attempts. *Journal of Consumer Research*, 21 (3), 540-550.